DOI 10.1007/s10958-017-3643-6
Journal of Mathematical Sciences, Vol. 228, No. 5, February, 2018

D-GUARANTEED DISCRIMINATION OF STATISTICAL HYPOTHESES: A REVIEW OF RESULTS AND UNSOLVED PROBLEMS

D. S. Simushkin[1] , S. V. Simushkin[1] , and I. N. Volodin[1]

We compare two sequential d-guaranteed tests and an optimal d-guaranteed test based on a fixed number of observations with respect to the average number of observations within the most accepted practical applications of Bayesian probabilistic models. We consider the sequential "first skipping" test, the sequential locally efficient test based on the score statistic, and the test based on a fixed number of observations that minimizes the necessary sample size. We study various characteristics of these tests connected with the number of observations within three probabilistic models, namely, the normal (ϑ, σ^2) distribution of the observed random variable and the normal a priori distribution of ϑ with fixed σ^2; the exponential distribution with the intensity parameter ϑ and the a priori gamma distribution of ϑ; and the Bernoulli sampling with the success probability ϑ and the a priori beta distribution of ϑ. We discuss the connection of the d-posterior approach with the compound decision problem as applied to the analysis of data provided by microchips (when the false discovery rate, or FDR, for short, is treated as the d-risk of the first kind). We present the vast data on characteristics of the mentioned tests obtained by the method of mathematical modeling in several tables. We discuss unsolved problems of the d-guaranteed discrimination of hypotheses with the minimal number of observations and approaches to their solution.

1. Introduction

The idea of the d-posterior approach to the guaranteed discrimination of hypotheses was inspired by the breakthrough in the Bayesian theory of statistical inference in the middle of the 20th Century. In the paper [10] J. Neumann described a wide class of statistical problems, where the output parameter θ was a realization of a random variable, and the problem consisted only in the specification of its a priori distribution. In such problems one usually possesses a large amount of data obtained earlier by researchers of analogous objects; these data give some information on the a priori distribution of the parameter. H. Robbins was one of those who led the development of the mentioned breakthrough. The advantage of his approach consists in the proposed technique for the estimation of the Bayesian risk on the basis of the analysis of archival data.

Note that prior to the works of H. Robbins, S. N. Bernstein [3] criticized the application of the standard (non-Bayesian) approach to the guaranteed statistical quality control problem. Constraints imposed on consumer's and producer's risks do not correspond to the essence of the problem; namely, the quality control has to monitor the rate of defective products received by a consumer, rather than the rate of nonstandard specimens erroneously treated as standard. The known approach ignores the actual a priori distribution of the controlled characteristic. However, even if the a priori distribution of the considered parameter is not known, in quality control problems there always exists a large body of data of previous inspections of analogous specimens; their analysis would allow one to propose the a priori distribution following the ideas of H. Robbins.

Such an approach to the notion of the statistical risk is specific for any statistical research implying the analysis of sequential observations like specimens such as data of demographic (for small regions) or health surveys and data provided by biochips in genetic research. In most of the mentioned problems the Bayesian risk is a very rough characteristic of average losses. One should average the losses only

[1] Kazan State University, Kazan, Russia, e-mail: `smshkn@gmail.com`

Proceedings of the XIX International Seminar on Stability Problems for Stochastic Models, Svetlogorsk, Russia, October 10–16, 2011